crystallization communications

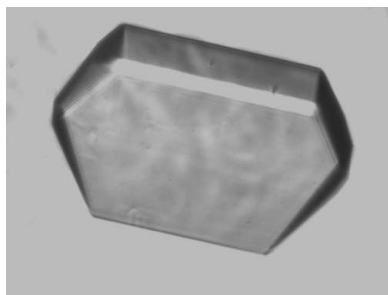# Overproduction, purification, crystallization and preliminary X-ray diffraction studies of the human spliceosomal protein TXNL4B

Tengchuan Jin,[a] Andrew J. Howard,[a] Erica A. Golemis,[b] Yingtong Wang[a] and Yu-Zhu Zhang[a]*

[a]Department of Biology, Illinois Institute of Technology, Chicago, IL 60616, USA, and [b]Fox Chase Cancer Center, Philadelphia, PA 19111, USA

Correspondence e-mail: yuzhu.zhang@iit.edu

The human gene coding for the spliceosomal protein thioredoxin-like 4B (TXNL4B) was overexpressed in *Escherichia coli* and the encoded protein was purified and crystallized. Well diffracting single crystals were obtained by the vapor-diffusion method in hanging drops. The crystals belong to the primitive monoclinic space group $P2$, with unit-cell parameters $a = 39.0$, $b = 63.6$, $c = 51.0$ Å, $\beta = 92.484°$, and diffract to at least 1.50 Å. A SeMet derivative of the protein was prepared and crystallized for MAD phasing.

1. Introduction

In eukaryotic cells, an important step in gene expression is the removal of introns from pre-messenger RNA. After recognition of the 5' and 3' end of the introns, the actual excision of the introns is catalyzed by the spliceosome, which contains four small nuclear ribonucleoprotein (snRNP) particles (U1, U2, U4/U6 and U5) and numerous auxiliary proteins (see, for example, Kramer, 1996; Zhou *et al.*, 2002). Besides the eight highly conserved common core proteins (B/B', D1, D2, D3, E, F and G Sm), snRNPs also contain highly conserved proteins specific for individual snRNP particles (see, for example, Will & Luhrmann, 1997), underlining the importance of these proteins. Before the structures of the spliceosome or the snRNP particles are obtained, structural information on individual snRNP proteins is essential to understand the many aspects of the pre-mRNA splicing process, including the structural rearrangements of the snRNPs. So far, much has been learned about the interaction and function of distinct components of the spliceosome by structure determination. These include structural studies of the complex of the U1A RNA-binding domain and part of the U1 snRNA (Howe *et al.*, 1994; Oubridge *et al.*, 1994), structural studies of the U2B''–U2A' complex associated with a fragment of U2 snRNA (Price *et al.*, 1998) and structural studies of the evolutionarily highly conserved U5 snRNP-specific protein TXNL4A (thioredoxin-like 4A, also known as U5-15kD and human dim1; Reuter *et al.*, 1999; Zhang *et al.*, 1999, 2003). However, little is known about the structure and interactions of other snRNP proteins.

Interestingly, during our studies of TXNL4A, we identified a previously undescribed putative protein, TXNL4B (a protein with 149 residues, also termed dim2 and DLP), that is homologous to TXNL4A (with one single amino-acid deletion in both proteins; the whole sequence of TXNL4A was aligned to the first 142 amino acids of TXNL4B with 38% identity; Zhang *et al.*, 1999). TXNL4A and its orthologs in other species associate with multiple proteins associated with pre-mRNA processing, including PQBP, hnRNP F, hnRNP H' and prp6 (Uetz *et al.*, 2000; Zhang *et al.*, 2000). To determine whether TXNL4B is functionally related to TXNL4A, we tested the ability of TXNL4B to interact with known TXNL4A-interacting proteins and found that TXNL4B was unable to interact with PQBP1, hnRNP F or hnRNP H' (data not shown). Recently, Sun and coworkers reported that TXNL4B associated with prp6 and blocking TXNL4B protein activity by RNAi led to insufficient pre-mRNA splicing (Sun *et al.*, 2004), suggesting that TXNL4B is a novel protein component of the spliceosome. We decided to initiate a structural study of TXNL4B, as determination of its three-dimensional structure will allow us to compare in detail any structural differences between TXNL4A and

TXNL4B and hence might reveal valuable insights into the functional differences between these two proteins. Here, we report details of the overexpression, purification and crystallization of human TXNL4B as well as preliminary crystallographic data obtained from TXNL4B crystals.

2. Methods, results and discussion

2.1. Cloning, expression and purification

The coding region of TXNL4B was PCR-amplified from a two-hybrid cDNA library. Two overlapping PCR fragments were produced to introduce a silent mutation in the internal *Nde*I site with overlapping primers. Another PCR reaction with a mixture of the above overlapping fragments as template produced the final PCR product, which was cloned into the pBluescript II SK vector *via* an *Eco*RI and an *Xho*I site to create pBlueTXNL4B. (*Eco*RI and *Nde*I sites were introduced by the 5′ primer and *Xho*I and *Bam*HI sites by the 3′ primer.) The TXNL4B-coding region was released from pBlueTXNL4B with *Nde*I and *Xho*I and cloned into the vector pET29b to create pET29TXNL4B. The insert of pET29TXNL4B was confirmed by sequencing. Induction of TXNL4B was revealed by an intense band at 17 kDa by SDS–PAGE and Coomassie staining.

Escherichia coli BL21(DE3) transformed with pET29TXNL4B were grown in 1 l LB medium containing 50 mg l^{-1} kanamycin at 310 K. At an OD_{600} of 1.2, expression of TXNL4B was induced by adding IPTG to a final concentration of 1 m*M*. The cells were collected by centrifugation after 4 h induction.

For purification, the cell pellet was resuspended in 50 ml buffer *A* (10 m*M* phosphate pH 7.9, 5 m*M* DTT) plus protease inhibitors (at final concentrations of 100 n*M* aprotinin, 50 µ*M* antipain, 50 µ*M* leupeptin and 0.5 µg ml^{-1} pepstatin). Cells were sonicated in a beaker surrounded by ice-cold water for 10 min. After centrifugation at 25 000*g* for 30 min at 277 K, an equal amount of saturated $(NH_4)_2SO_4$ was added to the supernatant and the mixture was centrifuged again at 25 000*g* for 30 min at 277 K. The supernatant was filtered and desalted with a Sephadex G-25 fine column. After desalting, the sample was loaded onto a Source 15Q ion-exchange column with a bed volume of 10 ml pre-equilibrated with buffer *A*. The column was washed with three bed volumes of buffer *A* and then run with buffer *A* and a linear NaCl gradient of 0–0.3 *M* (six bed volumes) at a flow rate of 1 ml min^{-1}. TXNL4B eluted at 0.1 *M* NaCl and the TXNL4B-containing fractions were concentrated to ~2 ml using an Ultrafree 5K centrifugal filter (Millipore, MA, USA). The concentrated protein was rapidly diluted to 80 ml with buffer *B* (10 m*M* acetate pH 4.46) and loaded onto a Source 15S ion-exchange column with a bed volume of 10 ml pre-equilibrated with buffer *B*.
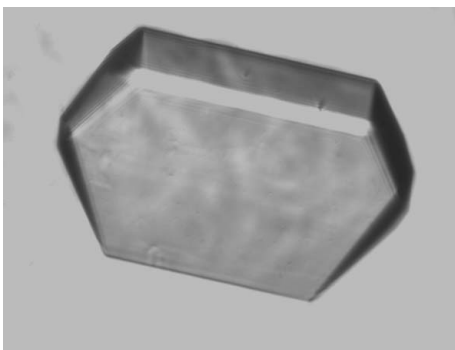
Table 1
X-ray data-collection statistics.

Values in parentheses are for the outer shell.

No. of crystals	1
Resolution (Å)	27–1.5 (1.59–1.5)
Wavelength (Å)	1.0
Space group	*P*2
Unit-cell parameters	
a (Å)	39.0
b (Å)	63.6
c (Å)	51.0
β (°)	92.4
Data-collection temperature (K)	110
Total No. of reflections	284263
No. of reflections	77634
Completeness (%)	99.1 (99.1)
$I/\sigma(I)$	19 (3.04)
R_{sym} (%)	5.07 (29.4)

The column was washed with three bed volumes of buffer *B* and then run with buffer *B* and a linear gradient of 0–0.5 *M* NaCl (six bed volumes) with a flow rate of 1 ml min^{-1}. TXNL4B eluted at 0.25 *M* NaCl and the TXNL4B-containing fractions were concentrated to ~4 ml using an Ultrafree 5K filter. The concentrated sample was fast diluted to 50 ml with buffer *A* and concentrated again before loading onto a Superdex75 gel-filtration column (bed diameter 26 mm, height 65 cm). The column was pre-equilibrated and eluted with buffer *A* at a flow rate of 1 ml min^{-1}. Pure TXNL4B eluted at 180 ml, with an apparent molecular weight close to that of a dimer as assessed by comparison with gel-filtration runs of standard proteins. All chromatographic steps were carried out at 277 K using an FPLC system (Pharmacia, Sweden). The molecular weight of the pure protein was determined to be 16 883 Da by electrospray mass spectrometry, consistent with the TXNL4B protein lacking the N-terminal methionine. The pure protein was concentrated and desalted using an Ultrafree 5K filter. The concentration of the final sample for crystallization was 50 mg ml^{-1} in Milli-Q H_2O.

2.2. Crystallization

The crystallization experiments were performed at 298 K in Linbro plates with the hanging-drop vapor-diffusion technique. Initial crystallization screening was performed using Hampton Research Crystal Screen and Crystal Screen Lite kits. 1 µl protein sample was mixed with 1 µl reservoir solution and sealed against 0.5 ml reservoir solution. All reservoir solutions contained 5 m*M* DTT and 0.02%(*w/v*) sodium azide. Numerous crystals appeared in wells containing solutions 6, 18 and 25 of the Crystal Screen kit after times ranging from several minutes to overnight. After optimization, single crystals of dimensions of ~0.6 × 0.4 × 0.15 mm were obtained within a week under the following conditions: 18% PEG 8000, 0.2 *M* magnesium acetate tetrahydrate, 0.1 *M* sodium cacodylate pH 6.5 (Fig. 1).

2.3. X-ray diffraction experiments and crystal characterization

To collect data at low temperature, single crystals were transferred with nylon loops to a cryoprotectant (1:1 mixture of reservoir solution and 40% sucrose). The crystals were then flash-frozen in nylon loops in liquid nitrogen. Data were collected at IMCA-CAT beamline 17-ID with an ADSC Quantum 210 detector at the Advanced Photon Source at Argonne National Laboratory. A complete data set of 0.5° frames with 4 s exposures was collected at 12 398 eV. The TXNL4B crystals diffracted to 1.5 Å. Data processing using the *HKL* suite of programs (Otwinowski & Minor, 1997) and *X-GEN* (Howard, 1997) revealed an monoclinic crystal system with unit-cell parameters *a* = 39.0, *b* = 63.6, *c* = 51.0 Å, *β* = 92.484°. From systematic absences

along the *b* axis, the space group could be determined to be *P*2 (Table 1). However, as size-exclusion chromatography indicates that TXNL4B is dimeric in solution, we think that there are two TXNL4B monomers in the crystallographic asymmetric unit. The Matthews coefficient is then $V_M = 1.85$ Å^3 Da^{-1}, with a solvent content of 34%.

2.4. Purification and crystallization of selenomethionine TXNL4B

Our first attempt to solve the structure of TXNL4B by molecular replacement using the structure of TXNL4A (PDB code 1qgv) as a model was unsuccessful. To facilitate multiwavelength anomalous diffraction (MAD) phasing, the met$^-$ *E. coli* strain B834(DE3) transformed with pET29TXNL4B was grown at 310 K in 1 l M9 medium containing 0.15 m*M* thiamine, 50 mg l^{-1} kanamycin and 50 mg L-selenomethionine. At OD$_{600}$ = 1.2, induction of SeMet TXNL4B expression was initiated with 1 m*M* IPTG. Expression was allowed to continue for 4 h. Purification of the SeMet TXNL4B was identical to that of the native protein. The success of the SeMet incorporation was verified by electrospray mass spectrometry. A difference in molecular weight of 140.4 Da between the natural and SeMet TXNL4B was measured that corresponded to full incorporation of three Se atoms (data not shown). SeMet TXNL4B crystals were grown under the same conditions as for native TXNL4B crystals. SeMet TXNL4B single crystals were also obtained within a week that were morphologically identical and grew to the same size as the native TXNL4B crystals. The SeMet TXNL4B crystals diffracted to 2.5 Å in an initial survey.

References

Howard, A. J. (1997). In *Crystallographic Computing 7*, edited by P. E. Bourne & K. D. Watenpaugh. Oxford University Press.

Howe, P. W., Nagai, K., Neuhaus, D. & Varani, G. (1994). *EMBO J.* **13**, 3873–3881.

Kramer, A. (1996). *Annu. Rev. Biochem.* **65**, 367–409.

Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.

Oubridge, C., Ito, N., Evans, P. R., Teo, C. H. & Nagai, K. (1994). *Nature (London)*, **372**, 432–438.

Price, S. R., Evans, P. R. & Nagai, K. (1998). *Nature (London)*, **394**, 645–650.

Reuter, K., Nottrott, S., Fabrizio, P., Luhrmann, R. & Ficner, R. (1999). *J. Mol. Biol.* **294**, 515–525.

Sun, X., Zhang, H., Wang, D., Ma, D., Shen, Y. & Shang, Y. (2004). *J. Biol. Chem.* **279**, 32839–32847.

Uetz, P., Giot, L., Cagney, G., Mansfield, T. A., Judson, R. S., Knight, J. R., Lockshon, D., Narayan, V., Srinivasan, M., Pochart, P., Qureshi-Emili, A., Li, Y., Godwin, B., Conover, D., Kalbfleisch, T., Vijayadamodar, G., Yang, M., Johnston, M., Fields, S. & Rothberg, J. M. (2000). *Nature (London)*, **403**, 623–627.

Will, C. L. & Luhrmann, R. (1997). *Curr. Opin. Cell Biol.* **9**, 320–328.

Zhang, Y. Z., Cheng, H., Gould, K. L., Golemis, E. A. & Roder, H. (2003). *Biochemistry*, **42**, 9609–9618.

Zhang, Y. Z., Gould, K. L., Dunbrack, R. L. Jr, Cheng, H., Roder, H. & Golemis, E. A. (1999). *Physiol. Genomics*, **1**, 109–118.

Zhang, Y. Z., Lindblom, T., Chang, A., Sudol, M., Sluder, A. E. & Golemis, E. A. (2000). *Gene*, **257**, 33–43.

Zhou, Z., Licklider, L. J., Gygi, S. P. & Reed, R. (2002). *Nature (London)*, **419**, 182–185.